ChinaXiv合作期刊 图书情报工作
第62卷 第23期 2018年12月

领域前沿识别方法综述*

■ 罗瑞[1,2] 许海云[1,3] 董坤[1,2]
[1] 中国科学院成都文献情报中心 成都 610041
[2] 中国科学院大学经济与管理学院图书情报与档案管理系 北京 100190 [3] 中国科学技术信息研究所 北京 100038

摘要：［目的/意义］通过比较分析不同的领域前沿识别方法，总结各类方法在识别与预测前沿上的优缺点，为未来研究前沿的识别与预测提出改进意见。［方法/过程］通过综述国内外研究前沿相关的文献，辨析研究前沿的相关概念。分析目前主要的研究前沿识别方法，相较于传统的识别方法，重点对变革性研究前沿的识别方法进行归纳。总结当前领域前沿识别方法存在的主要问题，并提出改进建议。［结果/结论］就概念而言，对"研究热点"、"新兴研究"和"研究前沿"的概念内涵在时间和创新程度两个维度上进行了区分，依据创新程度的不同，研究前沿可以分为"常规性研究前沿"和"变革性研究前沿"；就识别方法而言，不同方法都有其适用场景。未来需要深入挖掘"研究前沿"，尤其要关注到主题之间的语义关联和多源数据的多元关系融合，以及对"变革性研究前沿"的前期征兆进行特征挖掘，并构建相应的识别与预测方法。

关键词：领域前沿 研究前沿 变革性创新 新兴研究 主题识别
分类号：G301
DOI：10.13266/j.issn.0252-3116.2018.23.015

1 引言

当今世界正处于第六次科技革命前夜，全球科技创新呈现新的发展态势和特征，深刻认识并准确把握领域科技发展前沿及创新态势，对于创新战略前瞻部署、创新资源优化配置具有至关重要的作用。研究前沿是科技创新过程中最新、最有潜力、最具前瞻性和引领性的研究方向。有效识别领域发展前沿，可以对未来的研究趋势做出有效预判，从而将人力、物力和财力精准投入到最具战略研究价值的科技前沿。如何在早期识别出研究前沿的预警信号，如何在一项研究初露端倪的时候就能对其潜力进行预测，是一个极具挑战性但对制定科学政策和进行科研评价至关重要的问题。

本文选取Web of Science和中国知网(CNKI)作为检索平台，检索与"研究前沿"相关的国内外学术论文，主要关注管理学，科学学研究中有关前沿方法识别的高质量论文。最终得到中文文献494篇，英文文献588篇。检索日期为2018年8月8日。将所检索的文献按照发表时间形成发文分布的时间序列图(见图1)，可以看出"研究前沿"这一主题的发展大致分为：萌芽期、发展期和快速增长期，近年来步入了快速增长期。通过进一步梳理检索所得文献发现，目前已有不少关于研究前沿的识别和预测方法，但这些方法之间缺少系统深入地对比分析，尤其是缺少对重大、变革性创新领域的前沿识别方法的关注，尚未将变革性创新研究纳入领域前沿识别的方法体系中。

鉴于此，本文首先辨析研究前沿的相关概念，之后从专家定性判断和科学计量定量识别两个角度对比分析当前主要的研究前沿识别方法。重点梳理现有领域研究前沿识别中的计量分析方法，主要包括引文分析、知识单元分析、时间序列分析、多源数据分析、多维指标分析以及变革性研究前沿识别方法6个方面。在此基础上，比较分析各种方法的优缺点，总结当前领域前沿识别方法存在的主要问题，最后，提出改进研究前沿

* 本文系国家自然科学基金项目"基于科学-技术主题关联分析的创新演化路径识别方法研究"(项目编号:71704170)、中国博士后科学基金资助项目"面向多关系融合的知识创新路径的识别与预测方法研究"(项目编号:2016M590124)和中国科学院青年创新促进会(项目编号:2016159)研究成果之一。
作者简介：罗瑞(ORCID:0000-0002-5234-8901)，硕士研究生；许海云(ORCID:0000-0002-7453-3331)，副研究员，博士，通讯作者，E-mail:xuhy@clas.ac.cn；董坤(ORCID:0000-0001-8455-9204)，博士研究生。
收稿日期:2018-04-15 修回日期:2018-08-13 本文起止页码:119-131 本文责任编辑:易飞

识别方法的建议。



图 1 "研究前沿"相关主题发文趋势

2 概念辨析及特性

2.1 概念提出及发展

D. J. Price 于 1965 年首次提出研究前沿(research front)的概念,他认为某领域的研究前沿可以从常被科学家积极引用的文献中体现,因为这一小批论文发展活跃且迅速,能将曾经发表的文献同新发表的文献进行连接,促进了科学的发展,可以视为一种"growing tip"("生长锥")或"epidermal layer"("表皮层"),是区分科学与非科学的关键[1]。英文术语中的"research front"和"research frontier"均被译为研究前沿,但前者是利用计量学方法进行研究前沿的预测,属于先验评价,是一种期望结果;后者是经领域同行专家评判后确定的研究前沿,属于实际的观察结果,是真正意义上的领域前沿研究[2],努力缩小两者之间的偏差是情报分析追求的终极目标。

H. Small 和 B. C. Griffith 认为研究前沿指的是高交互性文献聚类的结果,因为这类论文簇展现了科学领域内的"高水平活动"。他们采用同被引聚类方法提炼出"研究前沿"[3]。O. Persson 通过引文耦合识别研究前沿,施引文献被视为研究前沿,被引文献被视为知识基础[4]。S. A. Morris 等对于研究前沿的认识与 O. Persson 相似,也是通过文献耦合方法进行研究前沿的识别,将研究前沿定义为引用一组固定和时间不变的基础文献的集合,将被引文献视为对应的研究前沿的知识基础。此外,S. A. Morris 等还认为研究前沿是不连续的,很可能随着科学家新开始一个问题或者新解决一个问题而产生或者消失[5]。王立学等将国外对研究前沿的定义归纳为 3 类,并分别对应 3 种不同的文献计量识别方法:将高被引文献视为研究前沿的多采用同被引分析,将施引文献视为研究前沿的多采用文献耦合分析,将突发或热点主题视为研究前沿的多采用共词分析方法[6]。这类有关研究前沿较为经典的观点都是从科技引文出发,关注的是较为活跃或受到更多关注的一批文章,对于研究前沿的界定受限于数据源。而在后续对研究前沿的理解上,陈超美等学者认为研究前沿是一组突现的动态概念和潜在的研究问题,强调其新趋势和突变的特点[7]。S. P. Upham 等则认为研究前沿是在科技领域中最为动态变化、吸引最多科学家目光的一类研究主题,是科学发现和社会关注的融合[8]。郑彦宁等指出研究前沿是针对特定领域和特定时间而言的,强调研究前沿的发展速度快,学术交流活跃[9]。冯佳认为研究前沿是一组具有较高学术关注度的最新研究主题,其内涵特征是具有"高关注度"和"新颖性"[10]。

通过中外学者对研究前沿概念的界定可以看出,领域前沿最主要的特征有两个:①时间上,产生于近期并延伸到未来;②创新程度上,蕴含较高创新价值的研究。值得注意的是,科技创新中的创新程度有着巨大区别,依据创新程度的高低又可以将研究前沿分为渐进性创新和变革性创新。两类科技创新的前沿识别信号和方法也有着较大差异。因此,本文分别从时间和创新程度两个维度辨析研究前沿及其相关概念的内涵差异,以期更加清晰地界定研究前沿的概念与内涵。与研究前沿相近的概念有研究热点(hot research)和新兴研究(emerging research)和突破性技术(radical innovation)等。

2.2 时间维度上的内涵差异

2.2.1 "研究热点"与"研究前沿" 研究热点是热度高的研究主题,当前有很多研究人员关注和研究。不少学者将热点主题视为发生频率高的一类主题,并可依托社交媒体进行热点事件的捕捉[11-12]。钟镇在探讨研究热点和研究前沿之间的关系时,认为前沿选题有很大概率成为下一阶段的热点选题[2],因此,研究热点主题部分来源于研究前沿。伴随着学科共同体的不断加入,研究前沿发展到一定阶段,有可能成为下一段时期的研究热点主题。研究热点主题重点在于受人"关注"更多,"热度"较高,相比于研究前沿在时间轴的出现上有一定的滞后性。从价值创新维度上看,"关注量"并不能作为"价值力"的判断标准,受到关注不代表此类研究最终能推动科学和社会发展。之所以有很多人关注,主要原因可能是某领域主题的研究价值逐步得到了确定,由此吸引了众多研究人员,但也不排除研究人员对于新兴领域的跟风导致。

罗瑞，许海云，董坤. 领域前沿识别方法综述[J]. 图书情报工作,2018,62(23):119－131.

2.2.2 “新兴研究”与“研究前沿” 新兴研究主题是当下新出现且呈现增长趋势的研究主题。在沃顿商学院于2000年出版的《沃顿论新兴技术管理》(中文2002版)中，将新兴技术的概念阐述为：新兴技术是建立在科学基础上的革新，它们可能创立一个新行业或者改变某个老行业，既包括产生于激进革新的间断性技术，又包括通过集中多个过去的独立研究成果而形成的更具创新性的技术[13]。郭涵宁认为新兴研究主题是在科学共同体内获得公认的、年轻并且呈迅速增长势态的有关科学或者技术问题的研究主题，而研究前沿可以在特定领域下随着时间不断变化。新兴研究主题强调的是当下，研究前沿是在每个观察的时间段内都有其对应的研究前沿[14]。笔者认为如果以观察的时间段去辨析则丧失了对概念本质的认识，因为新兴研究领域和领域研究前沿都有对时间的要求，即是相比于“观察期”是一种“新出现”并呈现“增长态势”或者相比于“观察期”是一种“前沿领域”。因此，这两个概念本质的不同在于新兴研究领域虽然呈现“年轻化”和“快增长性”趋势并不代表它在未来是有研究价值和前景的领域研究前沿。Rotolo等人则将具备新奇性、快增长性、连贯性、显著影响以及不确定性的技术划分为新兴技术[15]。因此，新兴研究领域是相比于观察期是一种“新出现”并呈现“增长态势”或者相比于“观察期”是一种“前沿领域”。

2.3 创新程度上的内涵差异

T. Kuhn根据科学发现创新性程度的大小不同，将科学研究分为常规科学研究与创新性研究两大类，即渐进性创新与变革性创新[16]。渐进性创新又称为常规科学研究，是以现有的科学发展作为基础进行创新，而变革性创新是指领域发展过程中的引起科学范式或技术范式变革的研究。变革性研究相关概念主要包括突破性技术、颠覆性技术、创造性破坏、技术突破或重大技术变革、不连续性创新[17]。

C. M. Christensen首先提出颠覆性技术(disruptive technology)这一概念，并将其视为取代现有的主流技术，带来新的产品和服务功能，对产业或市场格局产生破坏性、颠覆性影响的一类技术[18]。张金柱将突破性创新技术界定为：技术创新的方法、产品、设备、材料等技术主题发生不连续性变化，并引发性能的跃迁或功能的变化，最终导致市场、产品、服务、商业模式等发生不连续性变化[17]。颠覆性技术强调的是科学技术带来的实际效果，其特征与新兴技术类似，两者均强调“革新”和对“实践”带来的影响。突破性创新(radical innovation)概念与颠覆性创新内涵十分接近。不同的是，颠覆性创新技术对“技术创新”程度的要求并不明确，它的核心视角是市场细分和价值体系，并不一定伴随技术突破[19]，而突破性创新强调的是科学发展的过程中产生了巨大的变革，重点在于自身的“创新性”。其是否会在市场上产生立竿见影的效果还未知。

2.4 相关概念辨析

上述概念与“研究前沿”的内涵及外延既有差别又有联系。本文通过时间和创新性两个维度辨析“领域前沿”与“研究热点”、“新兴研究”的内涵关系。图2为本文相关概念内涵辨析示意图，是以O为坐标原点、横轴为时间维度、纵轴为创新性维度的二维坐标图。在时间轴上，t0为观察时间点，t0将整个时间轴划分为过去、现在和未来3个时间段。

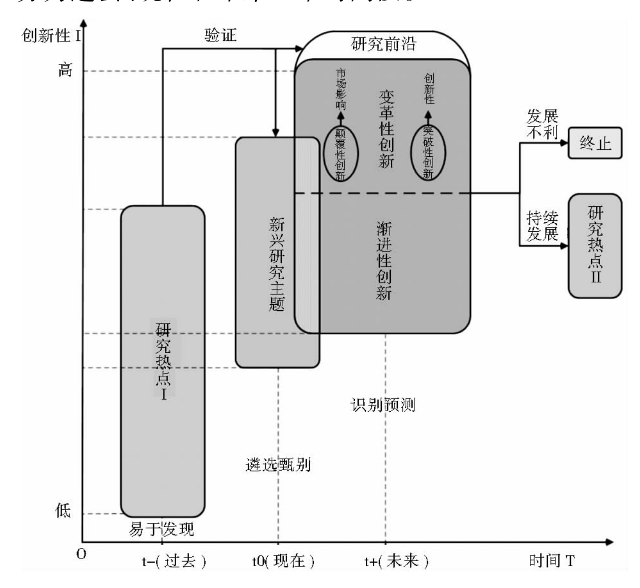


图2 “研究前沿”相关概念内涵辨析示意

总体而言，从时间维度上来看，“研究前沿”是更靠近未来，创新程度更高的一类创新主题，有更大的发展价值。缘于科技创新中的领域先占优势，最先对未知领域进行探索并取得创新的开拓者可以获得最大价值，因此，从创新程度上来看，相比于大家蜂拥而至的“研究热点”，“研究前沿”和“新兴研究”主题是创新性更高的一类主题。此外，对于研究前沿而言，根据创新程度又可分为创新度较低的“渐进性创新”和创新度较高的“变革性创新”。

具体而言，研究领域中的每个发展阶段都有其对应的“研究热点”，学者的关注包含对“研究前沿”的不完全预测，而以往“研究热点”中经过发展验证的部分进一步成为“新兴研究”和“研究前沿”。“研究热点”积累优势明显，主题特征显著，易于发现。相比于每个

阶段都会有对应的“研究热点”主题,“新兴研究”主题是指当下新出现的研究主题,是当下的和未来有发展潜力的一类主题,包括对“研究前沿”的探索,因此,与“研究前沿”存在交叉。由于“新兴研究”具有时间新颖性,通常没有“研究热点主题”的特征显著,需要利用特定的情报分析方法予以遴选甄别。同样,随着“研究前沿主题”在未来的持续发展也会吸引众多研究人员加入,最终可能成为“研究热点”,但也有可能因发展不利而终止。同时,“研究前沿”主题相比起“研究热点”和“新兴研究主题”,更侧重对未来高价值研究主题的探测,因此,需要更针对性的情报分析与预测方法识别“领域研究前沿”,通常识别和预测的难度也更大。

从以上与“研究前沿”相关的科学概念的辨析以及科学发现的分类中可以看出:“研究前沿”内涵丰富,具备多种特征,如“创新性”“变革性”“价值”和“发展潜力”等。“研究前沿”注重的是创新之后的科学价值,主要是指有发展前景的科学研究或者技术开发,注重的是主题或技术能对人类生活带来巨大改变的科技,无论是“渐进性创新”还是“变革性创新”,都蕴藏着极大的发展潜能和创新价值,这也是“研究前沿”最重要的特征。

由于目前对“研究前沿”没有统一的界定,研究人员对“研究前沿”的特征有不同的理解,因而对于“研究前沿”各种识别方法各有侧重。传统的科学计量方法对“渐进性创新前沿”有较好的识别效果,但“变革性创新前沿”充斥着随机性和突变性,相比于“渐进性创新”的有律可循,识别难度更大,识别方法也区别较大。下文对当前主要的研究前沿识别方法进行对比分析。

3 研究前沿识别方法分析

已有研究前沿识别方法大致可以归为两类:①专家判断法,即利用专家的知识对研究问题进行主观判断,依靠专家智慧综合判断得出最终结论;②计量分析方法,即通过对科技信息的量化分析来识别学科研究前沿。

3.1 专家判断法

目前,专家判断法是识别和预测科技发展态势的重要手段,不少行业领域和相关机构都有对领域前沿的预测分析,以把握领域发展方向。比较有代表性的产品有:《中国科学十大进展》、中国科学院《科学发展报告》、《Science》杂志评选的年度十大科学进展、《麻省理工科学评论》的十大突破性技术预测等,都是以专家智慧为主、数据分析结果为辅进行论证,分析得出科学技术发展现状和未来发展方向[17]。

专家判断可以充分利用专家的智慧和经验,但专家拥有自身独立心智模式,潜意识里更容易对自己感兴趣的领域给出更多关注,主观性强。M. Bengisu 和 R. Nekhili 在使用专利和文献计量方法对专家预测的突破性创新技术进行验证时发现,专家的主观判断可能产生巨大的主观偏差[20]。尤其是在研究对象动态变化性强、特征项难以提取的情况下,专家的智慧可能难以发挥最佳效能,不能及时快速地监测研究前沿,在速度、效率和准确性上没有太大优势。当前,随着数据密集型知识创新范式的出现和逐步深入,数据分析的作用日益凸显,其已成为专家判断中重要的辅助基础和方法。

3.2 基于引文的分析方法

基于引文的分析方法主要立足于文献与文献之间的引用关系,构建引用网络,通过对共被引知识网络的可视化分析识别研究前沿。单独使用该类方法进行研究前沿识别的研究较少,大多数研究是将该种方法与其他引文方法进行对比研究[21],共被引和耦合分析是领域前沿识别中更为常见的引文分析方法。两者的具体解读见表 1。

共被引分析方法存在时间滞后性,只有当一篇文章达到一定的被引次数后,才能得到关注,这往往会忽略潜在研究前沿。M. H. Huang 和 C. P. Chang 通过对比引文耦合和共被引分析方法在探测 OLED(有机发光二极管)领域的研究前沿时发现,引文耦合在检测的数量和速度上优于共引分析[28]。相比于共被引分析,耦合分析在一定程度上改进了时间滞后性的问题。其中引文耦合分析主要是针对施引文献展开,其参考文献是固定不变的,施引文献间的耦合关系可在施引文献发表后立即获得,且动态变化性没有共被引分析强,分析结果较为静态稳定[29]。但是如果是以作者等文献外部特征项作为耦合分析对象,耦合关系则会随着时间呈现动态变化,相比之下能从发展的角度更为全面地进行研究前沿的分析[30]。

除却以上两种较为常见的文献引用网络分析之外,D. Fajardo-Ortiz 等通过构建高互引论文的子网络来执行网络聚类的结合和文本挖掘来确定范式研究前沿并分析其动态[31]。但是这种方法在研究前沿的识别中的应用较少。

总而言之,文献之间的引用关系网络是识别和预

表 1　引文分析法对比说明

引文方法	图示	方法说明	代表性研究
共被引分析	研究前沿 II；研究前沿 I；A B C D E F G H I J	通常在构建文献共被引网络后，进行聚类分析，并结合对关键节点文献的内容分析，从而对研究前沿进行监测和识别	①许振亮和郭晓川[22] ②邓明君和罗文兵[23] ③潘黎和侯剑华[24] ④侯剑华等[25]
文献耦合分析	A B C D E F G H I J	通常在构建文献耦合网络后，进行聚类分析，并结合对关键节点文献的内容分析，从而对研究前沿进行监测和识别（耦合对象除了文献还可以是作者）	①I. Park 等[26] ②D. Zhao 和 A. Strotmann[27] ③S. A. Morris 等[5]

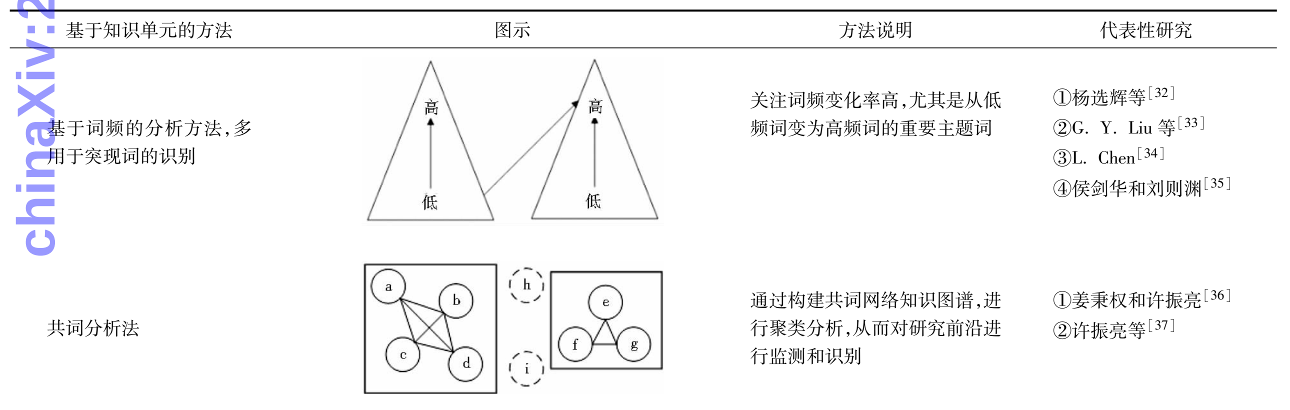
注：共被引分析法中，部分研究是从共被引聚类网络入手寻找"研究前沿"，还有部分研究是从共被引网络中的施引文献入手寻找"研究前沿"

测领域研究前沿的经典的重要方法。直接引用、共被引、引文耦合和互引等多种不同的引用方式都有着对研究前沿不同理解的体现，每一种引文网络均有其优劣势，在实际应用中，具体选择何种引文网络方法进行预测分析，需要从整个预测方法体系的构建中，依据不同引文分析网络的用处和功能而予以选择。

3.3　基于知识单元的领域前沿识别方法

基于知识单元的分析方法是从主题词的角度入手，对研究前沿进行识别。大多从两个角度出发。其一是从主题词的词频变化率入手，寻找突现词。其二是构建主题词共现网络，通过绘制共词知识网络图对研究前沿进行挖掘。两者的具体解读如表 2 所示：

表 2　基于知识单元的分析方法对比说明



基于知识单元的方法	图示	方法说明	代表性研究
基于词频的分析方法，多用于突现词的识别	高 低 高 低	关注词频变化率高，尤其是从低频词变为高频词的重要主题词	①杨选辉等[32] ②G. Y. Liu 等[33] ③L. Chen[34] ④侯剑华和刘则渊[35]
共词分析法	a b c d h i e f g	通过构建共词网络知识图谱，进行聚类分析，从而对研究前沿进行监测和识别	①姜秉权和许振亮[36] ②许振亮等[37]

词频分析法是众多定量分析方法中操作最为便捷的分析方法。该种方法主要是通过统计主题词的词频或者词频变化率来识别和监测研究前沿主题。杨选辉等在对"深阅读"主题进行新兴趋势分析时，将突变词划分为 4 类：成熟型突变词、衰退型突变词、突现型突变词和发展型突变词，并认为突现型突变词和发展型突变词能揭示该领域内的新兴趋势[32]。在对研究前沿进行识别时，大多数研究都是关注突现词。但仅使用词频分析或者突发词监测切断文献之间及主题词之间的语义关联，缺少连贯性，难以对研究前沿的知识结构进行分析。尤其是是在揭示领域知识结构时，并不是词频越高越好，根据齐普夫定律，对于词频阈值的选择，也是这类研究方法的难点。

共词分析弥补了词频分析方法中知识单元之间孤立的缺陷，将词与词之间进行有效连接，能在一定程度上展现知识结构的变化。此外，相比于引文分析，共词网络的结构不仅从微观层面揭示了科学知识体系内的实体关系特征，还以其演化过程反映了科学概念和科学命题的增长规律[38]。相比于词频分析法，共词分析构建了词与词之间的关系，能更好地挖掘知识单元间的结构关系，并且能处理的数据量也更大。这类方法的重点在于如何在共词网络中识别重要节点，提取有

价值的知识结构和知识社区。相比于引文网络,共词网络属于认知层面的知识网络,能从微观知识单元的角度对知识结构进行解读,知识单元也更细化、更准确。

3.4 基于时间序列的领域前沿识别方法

3.4.1 基于路径的领域前沿识别方法 科技创新不是凭空产生的,而是在现有研究内容的基础上交叉融合发展而来,因此,科技创新路径分析可以帮助更好地预测科技发展方向。科技创新路径是指将科技创新按照时间顺序衔接起来,勾勒出某项科技创新从首次出现到不断创新发展的演化过程[39]。通过路径演化趋势分析识别及预测科技前沿一直是情报分析的研究重点之一,相关研究最早可以追溯到 1965 年,"科学计量学之父"D. J. Price 在对研究前沿(research front)进行界定时,指出了科学发展路径中存在的"创新知识点"[1],体现出了基于引文链接的科技创新路径分析对于科技前沿识别及预测的重要作用。

结合引文链接与可视化图谱构建科技创新路径,并以之为基础,成为识别、预测科技前沿常见的方法。E. Garfield 利用直接引用网络生成了一个知识领域的历史演化图谱,并预测了该领域的发展趋势[40]。N. P. Hummon 等在 1989 年基于对引文分析方法的归纳总结,首次提出了一种关注引文网络中节点与节点之间的连接的分析方法,即主路径分析方法[41]。主路径分析方法对关键性文献的识别和主流研究线索的提取有重要意义[42]。随着文本挖掘技术的快速发展,基于主题词分析科技创新路径的发展趋势,预测科技前沿成为另一种主要方法。A. Kontostathis 等提出基于文本挖掘的自动探测方法 ETD (emerging trend detection),该方法首先将主题用一组时间特性关联的特征表示,然后根据这些特征用文本挖掘技术进行主题抽取,随着时间推移用一定的评价标准来关联主题,构建主题演化路径并判断其趋势,预测新兴趋势(科技前沿)[43]。这种知识演化路径的思路也被称为"数据流分析方法"。祝娜从主题关联角度入手,基于 LDA 模型识别科技创新主题并按照科研生命周期进行阶段细化构建了知识演化路径,以 3D 打印领域为例进行了实证研究,证明了基于时间序列的知识演化路径能够有效地探测科技前沿主题[39]。梁丽等采用数据流分析方法对学科的主题演化过程进行了解读[44]。

3.4.2 基于时间线的领域前沿识别方法 基于时间线的研究前沿识别方法能直观地展示研究前沿的动态变化。时间线技术是用 X 轴代表时间轴,Y 轴代表其他特征项,通过建立一个以时间为变量的研究主题函数,并将其反映在图谱中,使科学研究主题的变化脉络更加清晰[45]。基于时间线的分析方法是指在研究前沿识别过程中,通过对整个分析对象的时间线进行划分,来探讨研究前沿在整个生命周期中的演化过程。牵涉到研究对象如何随着时间的变化而变化时,均会用到该方法。例如,Gartner 公司在对突破性创新技术进行预测时,即将新科技的生命周期划分为 5 个阶段:萌芽期、引入期、成长期、成熟期和衰退期[17]。时间线分析是建立在识别研究主题的基础上,之后再进行研究前沿的确定或者演化趋势分析,归根结底仍要回到第一步的研究主题提取上,需要以一种识别研究主题的方法作为先行基础。

目前,该方法多用于研究前沿识别后的补充中,即在利用常用的研究前沿识别方法进行主题识别后,再通过时间线图谱分析呈现研究领域随时间的演化趋势[46]。除此之外,还有研究会直接在时间线图谱的演进分析基础上,结合对时区视图的分析对未来的研究趋势做出猜测[47-48]。张婷则在原来的时间序列研究前沿分析的基础上运用时间线和地形式可视化图谱分析方法,即由原来的二维时间线图谱扩展到三维动态地形式图谱,在可视化效果上进行了改善[49]。当前,随着信息技术的发展,研究人员开始综合利用可视化技术,深入文本内容进行主题演化分析,追踪研究前沿主题在时间维度上的分裂、融合等复杂演化过程,这些分析都能够在一定程度上弥补引文、文本内容分析方法的不足[50]。

3.5 基于多源数据的领域前沿识别方法

依据信息定量分析识别研究前沿的方法中,数据源的遴选是一项基础工作。数据源的可靠及充分与否会在较大程度上影响研究前沿的识别的效力。通常情况下单一数据源的信息量有限,整合利用多源数据也成为领域前沿识别中的重要工作。

从数据源的类型上来看,王贤文等将科学计量学的研究数据体系分为四大数据对象,分别是:发文数据、引用数据、使用数据和替代计量数据[51]。刘自强将研究前沿识别数据源分为:科技规划数据、基金项目数据、专利数据和论文数据[50]。在对研究前沿进行识别时,目前大多数研究都是集中在论文数据的发文数据和引用数据的分析上。也有研究对以上几类研究数据进行不同的组合,并综合利用多种方法进行研究前沿的识别。如 I. Park 等在对能源领域的研究前沿进行识别时,将专利文献和学术文献作为核心文献进行

罗瑞，许海云，董坤.领域前沿识别方法综述[J].图书情报工作,2018,62(23):119－131.

分析[26]。D. Zhao 和 A. Strotmann 比较了 XML 领域中期刊研究论文的作者共被引分析结果和网络研究论文的作者共被引分析结果,发现网络研究论文能更早地揭示领域研究的演进趋势[27]。

也有研究人员从不同数据源之间的知识关联关系入手,构建领域前沿的预测方法。白如江以科技规划文本和项目数据为研究对象,利用基于语义计算的科学研究前沿识别模型进行前沿识别[52]。孙震在综合了发文引用数据、下载使用数据、替代计量数据等多种数据后,结合引文分析、词频分析和共词分析等多种方法构建了研究前沿的集成识别模型[53]。许晓阳等也将论文与专利相结合,建立一套整合基础研究与应用研究的研究前沿识别方法[54]。张金柱综合专利数据和科学论文数据,从基础研究影响技术创新的角度出发,以专利引用科学论文为纽带研究突破性创新识别[17]。

但这些数据源的整合尚属于数据来源层的融合,集成程度和深度尚不够充分。多源数据融合可以按照融合阶段不同划分为 3 种类型:前期数据源与数据类型的融合,中期数据关系融合,后期聚类融合[55]。目前对于多源数据的融合大多还停留在前期的数据类型的融合和中期的数据关系的融合,对于后期的聚类融合的应用还不广泛。

3.6 多维指标领域前沿识别方法

因为领域前沿识别中单一指标的片面性,研究人员尝试利用多维指标识别和预测领域前沿。多维指标大致可以分为两种类型:一种是由于分析数据源的多样性而采用多重指标进行分析,一种主要是从研究前沿的特性出发,构建多维测算指标。后者可能是基于单一数据源,也可能基于多源数据源。表 3 列举了代表性的领域前沿识别多维指标方法。

表 3　领域前沿识别多维指标方法对比

研究人员	研究前沿属性		指标	优势	不足
郑彦宁等[9]	新颖性 时效性 集中性		研究主题年龄 研究主题关注作者数量	采用关键词共现方法,规避了引文分析的时间滞后及词频分析单一性的缺点	未将词语放置在句子语境中进行考量
范少萍[56]	新颖性		时效性指标 创新性	各项指标依据特性不同采用不同的分析对象,如参考文献、文本内容、专利和学科信息等,计算较为简单	指标主要是应用于科研基金申请方面,自然科学和人文科学有很大的不同,普适性不强,工作量大
	应用性		应用性指标		
	学科交叉性		学科交叉性		
	风险性		风险性指标		
张丽华[57]	研究前沿早期	时效性	时效性指数	对主题演化情况进行了详细的指标量化	综合前沿性指标受权重分配方案的影响,主观性强。同时,4 个分指标各有其适用范围,在不同领域的前沿探测过程中,需要根据领域发展的实际情况,单独或组合使用各项前沿特征指数
		突破性	突破性指数		
		跨学科性	跨学科性指数		
		继承性	继承性指数		
		综合	综合性指数		
	早期到进展期的演化规律		主题演化指数 主题演化率 主题演化强度率 前沿特征演化指数		
刘自强[50]	主题关注度		主题强度	综合考虑了影响主题演化分析的因素:时间、主题热度、主题状态和主题演化路径	在主题识别和主题演化可视化过程需要优化
	主题的关联、层次、分布和相互影响		主题结构		
	内部基本知识单元		主题内容		
R. J. Funk 和 J. Owen-Smith[58]	技术变革		CDt	抓住了发明在已有技术上的影响的方向	专利数据本身存在一定的局限性,如有些技术发明没有申请专利,有些公司可能为了战略等原因会规避引用行为
			mCDt	相比于 CDt,指标加入了影响程度的测量	
冯佳和张云秋[59]	主题强度		研究主题在全部科学文献中的权重的总和与总文献量的比值	基于 LDA 模型的指标整合方法结果准确且能处理大规模语料	对主题的解读依赖本体,现在各个领域的本体搭建尚不完善
	主题新颖度		主题的平均年龄		

3.7 变革性研究前沿识别方法

3.7.1 通过测度科学结构的变化识别研究前沿 C. Chen 等提出了科学知识不确定性的研究框架，利用科学图谱技术研究科学发展过程中的进化过程，发现科学知识的里程碑、关键路径、转折点和边界范围[60]。在对知识突破点进行识别时，在库恩的范式理论的基础上，认为科学革命是科学的重要组成部分，采用了渐进性的可视化方法，对 3 个方面做出了改进：改进单个网络的清晰度、凸显邻接网络的转变、识别潜在的节点。尤其对于潜在节点的识别，构建了路标节点(landmark node)、枢纽节点(hub node)和轴节点(pivot node)[61]。

3.7.2 通过分析弱关联关系识别研究前沿 弱相关关系(简称"弱关系")是学科交叉和技术融合的一个初期特征，也是预测突破性创新的一个重要切入点。弱关系一词最早起源于社会学，它与强相关关系(简称"强关系")是一个相互对立的概念，在网络结构中，弱关系指的是节点关系强度低于阈值的一类关系。M. Granovetter 认为强关系维系着组织内部的关系，而弱关系是群体、组织之间的信息传递纽带，它促成了不同群体之间的信息流动，传播了人们原本不太可能看到的信息，并且可以使一个较大的网络的结构更为凝聚[62]。弱关系不仅仅存在于人群群体之间，也存在于主题群体之间。研究前沿相比其他主题，更具前瞻性。这一特性使得其在知识发展的脉络中很可能受到忽视。就这个角度而言，弱关系在知识网络图谱中对于知识结构的形成发挥了重要的作用，在知识网络中也容易受到忽视，是研究前沿识别中需要关注的内容。

张英杰分别从整体网和个体网的角度探测网络中弱关系数据的演变情况，探测可能的发展前沿[63]。L. Wei 等结合弱关系分析，有效识别了情报学的交叉前沿[64]。目前对于变革性研究前沿的识别中，对于弱关系的关注还不多，在未来需要加强对知识弱关系进行有效信息的挖掘。

3.7.3 通过发掘睡美人文献识别研究前沿 变革性研究由于颠覆现有研究范式，科学共同体对此保持较大的心理距离，从而低估其知识价值，因此易被抵制[65-66]。科学和技术领域都有延迟承认现象，在科学中表现为睡美人文献，在技术中表现为睡美人专利(sleeping patents)[67]。睡美人文献和延迟承认现象的本质是科学研究的超前性或变革性，变革性研究前沿属于非线性发展的一类研究前沿，与睡美人文献的爆发前萌芽状态、引文曲线呈现前期无人引，后期高被引的发展态势有类似之处[68]，这也从另一角度说明睡美人文献是变革性研究前沿识别的一大来源，能为研究前沿的识别提供重要线索。因此，可以通过发掘睡美人文献识别研究前沿。

W. Glänzel 等基于平均数提出将"整体上超过 80% 的文献是发表 3 年内首次被引，超过 90% 的文献是发表后 5 年内首次被引"的这类文献界定为睡美人文献[69]。此外还有无参数客观指标的应用，如 J. Li 等将经济学中基尼(Gini)系数发展成为 G5 指数，运用到睡美人文献的识别过程中[70]。

3.8 领域前沿识别方法的比较研究

K. W. Boyack 和 R. Klavans 通过生物医学领域研究前沿识别的准确性上对 4 种引文方法进行了比较：共被引分析，引文耦合，直接引用和基于耦合的引文文本混合方法。结果发现：①从计算成本来看：前 3 种方法的计算成本大致相同，基于耦合的引文文本混合方法计算成本较高；②从分析结果聚类的覆盖面来看：共被引分析方法的覆盖面最大，混合分析方法和引文文献耦合方法紧随其后；③从连贯性和聚集性来看：在 3 种纯粹的引文分析方法中，文献耦合方法的连贯性和聚集性均优于其余三种方法，因此它是最为精确的研究前沿识别方法。但是混合方法在连贯性和聚集性上均对文献耦合方法进行了改进，是最优良的分析方法[21]。N. Shibata 等在研究中就共引网络、引文耦合和直接引文 3 种引文网络监测研究前沿的有效性上进行了比较。结果发现：直接引文在监测研究前沿方面展现了最好的性能，而共引分析性能最差[71]。这与 K. W. Boyack 和 R. Klavans 的研究结果不同。白如江等对比分析了目前科学研究前沿探测研究中主要使用的引文分析和主题词分析两种方法，指出目前存在 3 个方面的问题，即：引文分析存在时滞性；主题词分析缺乏语义信息支持；数据源无法有效融合[72]。

当前研究前沿识别领域有了较为成熟的理论和方法，如引文分析和主题词分析，被广泛应用于预测分析不同国家、学科的研究前沿，但是引文分析方法，存在引文时滞性的问题；主题词分析方法的分析内容相对孤立、欠缺关联分析，限制了研究前沿识别预测的及时性和准确性。表 4 为当前常用领域前沿识别方法的主要指标的性能和优缺点对比。

3.9 领域前沿识别方法存在的主要问题

3.9.1 创新路径中缺乏对前沿主题的深度关联分析 研究前沿的识别中的难点是知识新颖性的测度，涉及科学发现之前和之后的知识状态以及对科学发现的

罗瑞，许海云，董坤. 领域前沿识别方法综述[J]. 图书情报工作,2018,62(23):119－131.

表 4　领域前沿识别方法主要指标对比

分析方法	具体方法	具体指标				优点	缺点
		客观性	粒度	动态性	时效性		
专家判断法		×	√	√/×	√/×	建立在专家已有的较为完备、连接紧密的知识体系基础上，分析深入，操作便捷	会受到专家心智模型的影响，预测结果较为主观。在一般的前沿问题识别上，动态性和时效性较强，但是在面对复杂问题时，专家难以最快地对最新的研究问题做出预测
知识单元分析方法	词频统计	√	√	√	√	从主题词的角度入手，分析粒度更细且较为客观。相比于引文分析，动态性和时效性更强	主题词的确定十分复杂，没有对知识单元进行有效连接，语义关系弱
	共词分析	√	√	√	√	从主题词的角度入手，分析粒度更细且较为客观。相比于引文分析，动态性和时效性更强。此外有一定的主题关联	主题词的确定较为复杂，此外，对重点节点和社区网络的挖掘难度较大，语义关系挖掘难度大
引文分析方法	共被引	√	×	√	×	从被引文献出发，相比于引文耦合的动态变化性强	从被引文献出发，时间滞后，忽略了未被引用的文献，且分析粒度不及知识单元分析方法
	引文耦合	√	×	×	√	从施引文献出发，改善了共被引分析方法的时间滞后性	从施引文献出发，相较共被引分析，结果更为静态

注：就客观性而言，"√"表示客观性较强，"×"表示客观性较弱；就粒度而言，"√"表示粒度较细，"×"表示粒度较粗；就动态性而言，"√"表示动态性较强，"×"表示动态性较差；就时效性而言，"√"表示时效性较强，"×"表示时效性较差

新颖性和不可预测性的测度。因此，为了发现学科新兴趋势，必须首先了解学科领域中的科研主题的演化过程、规律及态势。主流的研究前沿识别方法，大多停留在对研究主题的识别，对研究主题的演化过程和动态变化规律的研究还不足。虽然已有部分研究关注到这一点，采用时间序列及涉及相关社区演化算法对主题演化进行分析，但是也仅仅停留在对于主题演化模式的描述性的规律总结，并没有对主题消亡的原因和兴起的原因做进一步的解读。

3.9.2　数据的单一化及多源数据融合的浅层化　已有前沿识别研究大多是基于科技论文和专利文献的引文或主题词进行分析，数据来源较为单薄。专利论文、基金数据、实验数据、科研论文、分析报告等均反映了科学发展的现状，并且蕴含了未来科学发展的趋势，当前尚未被充分运用。同时，常用的数据源大多存在时间滞后性和动态性不足的问题，每一种数据都存在其优势和劣势。当前也有部分研究尝试对多种数据源进行融合，多是将不同的数据源融入分析目标，但融合层次及深度多为前期融合。同时，伴随多源数据下的实体类型丰富多样，数据关系也日趋多元化，但当前大多研究前沿识别未考虑到多源数据下的多元关系融合，尚未对文本内容实体关系进行深层加工整合，因此，难以充分解释被分析实体的关联关系。

3.9.3　对变革性研究前沿的关注不足　当前的领域前沿识别方法更多关注渐进性创新的前沿识别，变革性创新前沿识别方法尚不多见。渐进性创新研究通常会遵循比较稳定的发展路线，而变革性创新的发展规律相比于渐进性创新研究的发展规律更难以获取，充满了不确定性和非线性特征，科学突破和创造性的发现等变革性研究在早期没有任何可以提前用来进行探测的迹象。由此，变革性研究前沿识别需要有别于渐进性前沿识别方法，这也意味着相关研究需要采用一些非常规性方法对这些不确定的科学主题进行识别。

4　领域前沿识别方法未来发展趋势

4.1　增强主题路径中前沿主题关联的研究

将主题放置在主题产生、发展的路径上是前沿主题识别的可行之路，用路径表示科技创新的过程可以展现创新演化过程中的创新主题的产生、发展、演变、融合或消亡等变化过程。在整个主题路径中可以对主题与主题之间的互动情况和主题的发展模式等规律进行总结，从而达到识别前沿主题的目的。在此过程中，伴随学科交叉主题的识别这一关键问题，在对学科交叉主题进行识别时，需要开展对学科交叉动力学机制的探讨，剖析学科交叉主题的成长规律，这样就能抽取出主题特征，进而设定相应的计量指标，达到识别交叉前沿的目的。

此外，在主题路径中对前沿主题进行识别时，还需要进行语义关联下的解读。在主题演化路径中，一个主题的发展变化会引起其他相关联主题的发展变化。因此我们不能以单一静态视角去看待一个主题的发展变化，应该回归到主题与主题之间的关联关系上动态地观察主题的演化规律。其中深层的关联即是语义关联。通过基于语义关联的创新演化路径分析识别领域

研究前沿,能在此基础上进一步识别研究前沿以及判断创新发展趋势。这也体现了只有通过加强对主题与主题之间关联的解读,才能全面准确地识别前沿。

当前,基于机器学习的实体语义关系的识别方法获得了极大发展,如何将其引入科技路径发展下的语义关联识别具有研究意义。当通过这种关联对科研主题之间关系进行判断时,只有结合特定的学科背景,深入分析主题之间的学科定位、学术目标、发展历史、研究方法、研究人员等多维度信息,基于多种维度下,全方位地对主题进行语义关联网络的构建,才能对前沿主题做出最准确的预测。

4.2 增强多源数据与多元关系融合分析

研究前沿识别需要融合该领域的各类信息进行综合判定。过去的研究中常用论文和专利数据,是因为受限于信息载体和科学技术。当代社会,信息技术的快速发展催生了多种信息载体。这些信息源都会在各个方面或多或少地影响领域的发展,同时也折射了领域发展的前景。但在多源数据的收集中,还需要明晰的是:收集有用的信息,去除无用的信息。

未来前沿主题识别的数据分析方法需要增强多源数据中期与后期融合,由此挖掘出更多主题关联信息。信息之间的关系结构是科学知识结构的映射,只有明晰了科学知识结构,才能对研究前沿做出合理的判断。引用关系、耦合关系和共现关系等多种数据关系都是知识发展脉络的体现,如何从已有的多种数据关系中去挖掘知识发展动向是未来的突破点。这类方法的难点在于如何整合不同的数据源、如何整合不同的数据关系及如何对多元关系融合后的综合型矩阵进行聚类。同时,随着大数据分析方法的不断完善,寻求领域前沿识别中,对多源动态数据的高效深入分析,预期会推动前沿识别的时效和精准度迈上新台阶。

4.3 捕捉变革性研究中不确定性

现有的研究前沿探测方法大多是在研究领域发展的进展期或成熟期进行前沿探测,此时大量研究人员已经进入该领域,探测的结果前瞻性已经不强[57]。变革性创新需要较长的积累时间,识别阶段越靠近前端越好,更利于进行技术实施和产业布局,但阶段越早,信息量越少,难度越大。许多科学突破和创造性的发现,其实没有任何可以让人们提前发现或利用的早期迹象。有关早期迹象的问题,以及如何在研究项目的开始时期测量变革潜力的问题,是极具挑战性的问题。深入挖掘变革性创新的初期特征是准确识别变革性研究前沿的前提。科技创新过程往往伴随着不确定性,变革性研究作为一种非线性创新,尤其受到这种不确定性的影响,因此在对变革性研究前沿进行识别时,把握这种会影响深远的不确定性尤为关键。我们需要深入挖掘变革性创新的初期特征,对这些初期征兆设计相应的甄别指标,从而能对变革性研究进行遴选甄别。

未来,从科学的不确定性角度出发进行领域研究前沿的识别,是从更为微观、基础的层面理解和认知变革性研究的特征。在对不确定性信息进行捕捉时,值得关注的是弱关系。前沿主题在发展萌芽之初甚至之前,往往以学科间的微弱关系形式存在,因此弱关系可以视为探索变革性前沿初期特征的一个切入点,它是数据挖掘中不可忽视的重要分析对象,代表了与高频词和强关系不同的数据对象,弱关系数据可能反映事物真实的性质,对该类型数据的忽略可能导致重要信息的遗漏。未来可以尝试通过捕捉变革性研究过程中类似弱关系这种代表科学发展的不确定性的信息来帮助识别研究前沿。

5 结语

本文通过对研究前沿相关主题的科学文献的调研,系统比较分析了目前已有的相关识别方法。从调研结果来看,领域前沿的研究已从多角度展开,但缘于研究前沿具有前瞻性和动态变化性,尚无哪种方法能通过定量分析有效识别领域前沿。当前情报分析中的领域研究前沿更多是一种预判,是否会成为真正的前沿主题,尚待验证。而实践中的研究前沿通常经过领域专家判断和认可,更接近于真实情况下的研究前沿。未来,随着可分析信息的不断增多与分析方法的不断改进,努力缩小两者之间的偏差是情报分析追求的终极目标。同时,深入挖掘研究前沿的特征是构建精准识别领域前沿的关键,但目前对研究前沿的特性和计量特征挖掘还不足。另外,已有的识别方法大多针对渐进性创新前沿主题,相比之下变革性研究前沿的识别面临更大的挑战。最后,本文提出未来可以通过增强主题路径中前沿主题关联的研究,增强多源数据与多元关系融合和捕捉变革性研究中不确定性与弱关系等途径来完善领域前沿的定量识别方法。

参考文献:

[1] PRICE D J. Networks of scientific papers[J]. Science, 1965, 149(3683):510-515.

[2] 钟镇. 从高被引与零被引论文的引文结构差异看 Research Front 与 Research Frontier 的区别[J]. 图书情报工作, 2015, 59(8):87－96.

[3] SMALL H, GRIFFITH B C. The structure of scientific literatures I: identifying and graphing specialties[J]. Science studies, 1974, 4(1):17－40.

[4] PERSSON O. The intellectual base and research fronts of JASIS 1986－1990[J]. Journal of the American Society for Information Science, 1994, 45(1):31－38.

[5] MORRIS S A, YEN G, WU Z, et al. Time line visualization of research fronts[J]. Journal of the Association for Information Science & Technology, 2003, 54(5):413－422.

[6] 王立学, 冷伏海. 简论研究前沿及其文献计量识别方法[J]. 情报理论与实践, 2010, 33(3):54－58.

[7] 陈超美, 陈悦, 侯剑华, 等. CiteSpace Ⅱ:科学文献中新趋势与新动态的识别与可视化[J]. 情报学报, 2009, 28(3):401－421.

[8] UPHAM S P, SMALL H. Emerging research fronts in science and technology: patterns of new knowledge development[J]. Scientometrics, 2010, 83(1):15－38.

[9] 郑彦宁, 许晓阳, 刘志辉. 基于关键词共现的研究前沿识别方法研究[J]. 图书情报工作, 2016, 60(4):85－92.

[10] 冯佳. 研究前沿识别与分析方法研究[D]. 长春:吉林大学, 2017.

[11] ISHIKAWA S, ARAKAWA Y, TAGASHIRA S, et al. Hot topic detection in local areas using Twitter and Wikipedia[C]// ARCS workshops. Muenchen:IEEE, 2012:1－5.

[12] KIM H G, LEE S, KYEONG S. Discovering hot topics using Twitter streaming data: social topic detection and geographic clustering[C]// IEEE/ACM international conference on advances in social networks analysis and mining. Niagara Falls:IEEE, 2014:1215－1220.

[13] 休梅克. 沃顿论新兴技术管理[M]. 石莹,等译. 北京:华夏出版社,2002.

[14] 郭涵宁. 多元科学指标视角下的新兴研究领域识别探索[D]. 大连:大连理工大学, 2013.

[15] ROTOLO D, HICKS D, MARTIN B R. What is an emerging technology? [J]. Research policy, 2015, 44(10):1827－1843.

[16] KUHN T. The structure of scientific revolution [M]. Chicago: University of Chicago Press, 1999.

[17] 张金柱. 利用被引科学知识突变识别突破性创新[M]. 北京:科学出版社,2017.

[18] CHRISTENSEN C M. The innovator's dilemma[M]. Boston: Harvard Business School Press, 1997.

[19] 孙启贵, 邓欣, 徐飞. 破坏性创新的概念界定与模型构建[J]. 科技管理研究, 2006, 26(8):175－178.

[20] BENGISU M, NEKHILI R. Forecasting emerging technologies with the aid of science and technology databases[J]. Technological forecasting & social change, 2006, 73(7):835－844.

[21] BOYACK K W, KLAVANS R. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? [J]. Journal of the American Society for Information Science & Technology, 2010, 61(12): 2389－2404.

[22] 许振亮, 郭晓川. 国际技术创新研究前沿的科学计量学分析[J]. 图书情报工作, 2011, 55(8):49－53.

[23] 邓明君, 罗文兵. 国际碳足迹研究前沿演进的可视化[J]. 湖南科技大学学报(社会科学版), 2013, 16(3):94－98.

[24] 潘黎, 侯剑华. 国际高等教育研究的热点主题和研究前沿——基于8种SSCI高等教育学期刊2000－2011年文献共被引网络图谱的分析[J]. 教育研究, 2012(6):136－143.

[25] 侯剑华, 李莲姬, 杨秀财. 基于引文网络结构变换的大数据研究前沿预测[J]. 情报科学, 2018, 36(6):142－148.

[26] PARK I, LEE K, YOON B. Exploring promising research frontiers based on knowledge maps in the solar cell technology field[J]. Sustainability, 2015, 7(10):13660.

[27] ZHAO D, STROTMANN A. Can citation analysis of Web publications better detect research fronts? [J]. Journal of the American Society for Information Science & Technology, 2007, 58(9):1285－1302.

[28] HUANG M H, CHANG C P. A comparative study on detecting research fronts in the organic light-emitting diode (OLED) field using bibliographic coupling and co-citation[J]. Scientometrics, 2015, 102(3):2041－2057.

[29] 孙涛涛, 杨立英. 基于文献耦合的研究前沿监测方法与应用研究[C]//中国医学科学院. 中国医学科学院/北京协和医学院医学信息研究所/图书馆2010年学术年会论文集. 北京:北京协和医学院/中国医学科学院,2011:564－568.

[30] 陈远, 王菲菲. 基于CSSCI的国内情报学领域作者文献耦合分析[J]. 情报资料工作, 2011(5):6－12.

[31] FAJARDO-ORTIZ D, LOPEZ-CERVANTES M, DURAN L, et al. The emergence and evolution of the research fronts in HIV/AIDS research[J]. Plos One, 2017, 12(5):e0178293.

[32] 杨选辉,杜心雨,蔡志强. 基于突变检测与共词分析的深阅读新兴趋势分析[J]. 图书馆建设,2018(5):48－53.

[33] LIU G Y, JIANG R, JIN Y. Sciatic nerve injury repair: a visualized analysis of research fronts and development trends[J]. Neural regeneration research, 2014, 9(18):1716－1722.

[34] CHEN L. Information visualization analysis of the hot research topics and the research fronts in the field of public information resources management (PIRM) based on Web of Science[C]// 2009 international conference on public administration. Chengdu: UESTC Press, 2009:944－948.

[35] 侯剑华, 刘则渊. 纳米技术研究前沿及其演化的可视化分析[J]. 科学学与科学技术管理, 2009, 30(5):23－30.

[36] 姜秉权, 许振亮. 基于知识图谱的国际生态文明研究前沿:共

词分析视角[J]. 科技与经济, 2009, 22(5):55－58.
[37] 许振亮, 赵宇娜, 周文霞. 国际生态城市研究前沿的知识图谱分析——基于共词网络分析视角[J]. 城市发展研究, 2013, 20(2):115－124.
[38] 王晓光. 科学知识网络的形成与演化(I):共词网络方法的提出[J]. 情报学报, 2009, 28(4):599－605.
[39] 祝娜. 语义增强的科技创新路径识别研究[D]. 淄博:山东理工大学, 2015.
[40] GARFIELD E. Historiographic mapping of knowledge domains literature[J]. Journal of information science, 2004, 30(2): 119－145.
[41] HUMMON N P, DEREIAN P. Connectivity in a citation network: the development of DNA theory[J]. Social networks, 1989, 11(1):39－63.
[42] 韩毅, 金碧辉. 引文网络主路径分析方法的形成与演化[C]// 中国科学学与科技政策研究会. 第六届中国科技政策与管理学术年会论文集. 北京:中国科学学与科技政策研究会, 2010: 155－164.
[43] KONTOSTATHIS A, GALITSKY L M, POTTENGER W M, et al. A survey of emerging trend detection in textual data mining[M]. New York: Springer, 2004.
[44] 梁丽, 谢凤杰, 池丽旭, 等. 特定学科热点和前沿主题研究方法实证分析[J]. 图书馆杂志, 2018(1):19－26.
[45] 张婷. 科学传播研究的可视化分析[D]. 大连:大连理工大学, 2009.
[46] 于小艳, 卢晓中. 潘懋元教育研究发展脉络管窥——基于学术论文的知识图谱分析[J]. 山东高等教育, 2015(10):83－96.
[47] 王丹, 宫晶晶. 基于知识图谱的国内建筑安全领域可视化研究[J]. 工程管理学报, 2016, 30(6):43－48.
[48] 兰国帅. 国际教育技术研究前沿热点知识图谱建构研究——基于十八种 SSCI 期刊 1960－2016 年文献的可视化分析[J]. 现代远距离教育, 2017(3):57－76.
[49] 张婷. 时间线和地形式可视化图谱:科学传播研究前沿演进趋势分析[J]. 情报学报, 2009, 28(6):923－928.
[50] 刘自强. 基于 NSF 数据的研究前沿主题识别及演化规律研究[D]. 淄博:山东理工大学, 2017.
[51] 王贤文, 方志超, 胡志刚. 科学论文的科学计量分析:数据、方法与用途的整合框架[J]. 图书情报工作, 2015, 59(16):74－82.
[52] 白如江. 基于语义计算的科学研究前沿识别研究[D]. 北京:中国科学院大学, 2015.
[53] 孙震. 基于科学论文多源数据的研究前沿集成识别模型研究[J]. 情报杂志, 2016, 35(8):95－100.
[54] 许晓阳, 郑彦宁, 刘志辉. 论文和专利相结合的研究前沿识别方法研究[J]. 图书情报工作, 2016, 60(24):97－106.
[55] 许海云, 董坤, 隗玲, 等. 科学计量中多源数据融合方法研究述评[J]. 情报学报, 2018, 37(3):318－328.
[56] 范少萍. 基于文献计量学的研究前沿指标体系及其对我国科学基金评审的借鉴意义[J]. 中国科学基金, 2016(2):171－178.
[57] 张丽华. 研究前沿探测及其演化分析方法与实证研究[D]. 北京:中国科学院大学, 2015.
[58] FUNK R J, OWEN-SMITH J. A dynamic network measure of technological change[J]. Management science, 2017, 63(3):791－817.
[59] 冯佳, 张云秋. 基于 LDA 和本体的科学前沿识别与分析方法研究[J]. 情报理论与实践, 2017, 40(8):49－54.
[60] CHEN C, SONG M. Representing scientific knowledge: the role of uncertainty[M]. Cham: Springer International Publishing AG, 2018.
[61] CHEN C. Searching for intellectual turning points: progressive knowledge domain visualization[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(Sl):5303－5310.
[62] GRANOVETTER M. The strength of weak ties: a network theory revisited[J]. Sociological theory, 1983, 1:201－233.
[63] 张英杰. 科技领域前沿计量探测方法研究[D]. 北京:中国科学院文献情报中心, 2011.
[64] WEI L, XU H, WANG Z, et al. Topic detection based on weak tie analysis: a case study of LIS research[J]. Journal of data and information science, 2016, 1(4):81－101.
[65] CAMPANARIO J M. Rejecting and resisting Nobel class discoveries: accounts by Nobel Laureates[J]. Scientometrics, 2009, 81(2): 549－565.
[66] FANG H. An explanation of resisted discoveries based on construal-level theory[J]. Science & engineering ethics, 2015, 21(1): 41－50.
[67] PALOMERAS N. Sleeping patents: any reason to wake up? [J]. Iese research papers, 2003, 20(35):D506.
[68] 杜建. "睡美人"文献的识别方法与唤醒机制研究[D]. 南京:南京大学, 2017.
[69] GLANZEL W, SCHLEMMER B, THIJS B. Better late than never? on the chance to become highly cited only beyond the standard bibliometric time horizon[J]. Scientometrics, 2003, 58(3):571－586.
[70] LI J, SHI D, ZHAO S X, et, al. A study of the "heartbeat spectra" for "sleeping beauties"[J]. Journal of informetrics, 2014, 8(3):493－502.
[71] SHIBATA N, KAJIKAWA Y, TAKEDA Y, et al. Comparative study on methods of detecting research fronts using different types of citation[J]. Journal of the Association for Information Science & Technology, 2009, 60(3):571－580.
[72] 白如江, 冷伏海, 廖君华. 科学研究前沿探测主要方法比较与发展趋势研究[J]. 情报理论与实践, 2017(5):33－38.

罗瑞, 许海云, 董坤. 领域前沿识别方法综述[J]. 图书情报工作,2018,62(23):119－131.

作者贡献说明:
罗瑞:负责资料收集整理、论文撰写与修改;
许海云:指导论文的组织、撰写并指导修改;
董坤:指导论文撰写。

A Review of the Main Recognition Methods of Frontier Research

Luo Rui[1,2] Xu Haiyun[1,3] Dong Kun[1,2]

[1] Chengdu Documentation and Information Center, Chinese Academy of Sciences, Chengdu 610041

[2] School of Economic and Management, University of Chinese Academy of Sciences, Beijing 100190

[3] Institute of Scientific and Technical Information of China, Beijing 100038

Abstract: [**Purpose/significance**] This paper summarized the advantages and disadvantages of various recognition methods of frontier researches by comparing and analyzing these methods, and suggested improvements for future researches. [**Method/process**] Firstly, it reviewed the literatures related to research frontiers at home and abroad, and identified relevant concepts of the research frontier. Then, the paper analyzed the main exsited research methods for the identification of research frontiers. Comparing with the traditional identification methods, this paper emphasized the summary of recognition methods of transformative research frontiers. Finally, this study summarized the main problems existed in the frontier identification methods, and put forward suggestions for improving the research frontier identification methods. [**Result/conclusion**] In terms of conceptual meanings, this paper distinguished research hotspots, emerging researches and research frontiers in the dimensions of time and innovation. According to the different levels of innovation, research frontiers can be divided into conventional research frontiers and transformative research frontiers. In terms of identification methods, different methods have their own applicable scenes. In the future, it is necessary to further dig out the measurement features of the "research frontier", especially the semantic associations among topics and the fusion of multiple relations should be focused on, as well as the feature mining of the early signs of "transformation research frontier", and then constructing corresponding recognition and prediction methods.

Keywords: forefronts research frontier transformative innovation emerging research topic recognition